A New Post-processing Method to Deal with the Rotational Indeterminacy Problem in MCMC Estimation

Kensuke Okada[1] Shin-ichi Mayekawa[2]

[1]Senshu University

[2]Tokyo Institute of Technology

August 26 2010

Rotational indeterminacy

- Infinite number of resultant matrices account equally for an observed data.
- If $\mathbf{X}$ is a solution, then so is any isometric transformation of $\mathbf{X}$.
- When we represent the isometric transformation by $f(\cdot)$, the transformed configuration,

$$\mathbf{X}^* = f(\mathbf{X}),$$

is also a solution.

Rotational indeterminacy in MCMC

- In classical estimation, rotational indeterminacy is just a problem of rotating a single solution matrix.
- However, in MCMC each of the (thousands of) MCMC samples has the freedom of rotation etc.
    - Situation is more complex in MCMC.

Rotational indeterminacy in MCMC

- In classical estimation, rotational indeterminacy is just a problem of rotating a single solution matrix.
- However, in MCMC each of the (thousands of) MCMC samples has the freedom of rotation etc.
  - Situation is more complex in MCMC.
- Objective:
  - To propose a new method of dealing with rotational indeterminacy in MCMC.
  - To empirically compare it with existing methods by simulation study.

Existing method A: Use informative priors on $\mathbf{X}$

- One of the benefits of Bayesian analysis.
- Used in many studies, e.g.,
  - DeSarbo, Kim, Wedel & Fong (1998, *Europ J Oper Res*).
  - DeSarbo, Kim & Fong (1999, *J Econometrics*).
- However,
  - subject to criticisms for its subjectivity.
  - prior information may not always be available.

Existing method B: Fix some elements of $\mathbf{X}$ to be 0

- Reduces degree of freedom.
- Used in Bayesian analysis as well as classical analysis.
- Used in many studies, e.g.,
  - Wedel & DeSarbo (1996, *J Bus Econ Stat*).
  - Lopes & West (2004, *Stat Sinica*).
- However, it is often difficult to decide which element should be fixed.

Existing method #1: Eigen analysis

- At each MCMC iteration,
  - Centralize $\mathbf{X}^{(l)}$.
  - Rotate it by $\mathbf{x}_i^{*(l)} = \mathbf{Q}^{(l)\prime}\mathbf{x}_i^{(l)}$, where
    - $\mathbf{x}_i^{(l)}$ is the i-th row of $\mathbf{X}^{(l)}$.
    - $\mathbf{Q}^{(l)}$ is the matrix whose columns are the eigenvectors of the covariance matrix $\mathbf{S}_x^{(l)} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i^{(l)} - \bar{\mathbf{x}}^{(l)})'(\mathbf{x}_i^{(l)} - \bar{\mathbf{x}}^{(l)})$.
- Then use approximate posterior mode of $\mathbf{X}^*$ as an point estimate.
- Used by Oh & Raftery (2001, *JASA*)'s Bayesian MDS.

Existing method #2 / #3: Procrustes Analysis (on-line / barch)

- Rotate each $\mathbf{X}^{(l)}$ for a target matrix $\mathbf{X}_0$ by Procrustes rotation:
  $\mathbf{X}^{*(l)} = \text{arg min tr}(\mathbf{X}_0 - \mathbf{Q}^{(l)}\mathbf{X}^{(l)})'(\mathbf{X}_0 - \mathbf{Q}^{(l)}\mathbf{X}^{(l)})$.
  - $\mathbf{Q}^{(l)}$ ranges over the set of rotations, reflections, and transformations.
  - $\mathbf{X}_0$: (e.g.) classical MDS solution.
- Both of the followings processings are possible:
  - On-line: rotate at each iteration l.
  - Batch: rotate after whole sampling process.
- Used e.g. by Hoff et al. (2002, *JASA*).

Proposed method: Batch generalized Procrustes analysis

- Stephens (1997, *JRSS B*) proposed an idea to deal with label-switching problem in mixture models.
  - Post-process MCMC samples so that marginal posterior distributions of the parameters are unimodal and close to normal.
- We apply this idea to rotational indeterminacy problem.
- We denote l-th centered and normalized MCMC samples by $\mathbf{X}^{(l)}$.

Proposed method (cont'd)

- We rotate:

$$\mathbf{X}^{*(l)} = \mathbf{X}^{(l)}\mathbf{Q}^{(l)}$$

  where $\mathbf{Q}^{(l)}$ is the transformation matrix that minimizes

$$||\mathbf{X}^{(l)}\mathbf{Q}^{(l)} - \bar{\mathbf{X}}^{*}||. \tag{1}$$

  toghether with $\bar{\mathbf{X}}^{*}$(where $\mathbf{X}^{*(l)'}\mathbf{X}^{*(l)}$: diag).
- This minimization problem is solved by using generalized Procrustes rotation (Schönemann & Carroll, 1970, *Psychometrika*).
  - Alternating least squares algorithm is used to minimize (1).

Proposed method (cont'd)

1. (1) is consecutively minimized for $l = 1, ..., L$.
2. $\bar{\mathbf{X}}^*$ is updated after each step.

- The proposed criterion is equivalent to maximizing the likelihood of normal distribution,

$$L = \sum_i \sum_k \sum_l \frac{1}{\sigma} \exp\left(-\frac{1}{2} \frac{(x_{ik}^{*(l)} - \mu_{ik})^2}{\sigma^2} \right)$$

- This method does not require external target matrix such as $\mathbf{X}_0$ in Method #2, #3.

Simulation study: Compared methods

- We consider Bayesian MDS model (Oh & Raftery, 2001).
- Following four methods are compared:

1. Eigen analysis (original method).
2. On-line rotation to the target matrix (classical MDS solution).
3. Batch rotation to the target matrix (classical MDS solution).
4. Batch generalized Procrustes rotation [proposed method].

Bayesian MDS: Model

- $\mathbf{\Delta} = \{\delta_{ij}\}$: $(n \times n)$ Observed dissimilarity matrix
- $\mathbf{D} = \{d_{ij}\}$: $(n \times n)$ Distance matrix
- $\mathbf{X} = \{x_{ik}\}$: $(n \times p)$ Configuration matrix
- The observed dissimilarity δ_{ij} follows the truncated normal distribution,

$$\delta_{ij} \sim N(d_{ij}, \phi^2) I(\delta_{ij} > 0),$$

where

$$d_{ij} = \sqrt{\sum_k (x_{ik} - x_{jk})^2}.$$

Bayesian MDS: Priors

- For prior of $\mathbf{x}_i$, a multivariate normal distribution is used:

$$\mathbf{x}_i \sim N(\mathbf{0}, \boldsymbol{\Lambda}). \quad (i = 1, ..., n)$$

- For prior of ϕ^2, an inverse gamma distribution is used:

$$\phi^2 \sim IG(a, b).$$

- For the elements of $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_p)$, an inverse gamma distribution is used:

$$\lambda_k \sim IG(\alpha, \beta_k). \quad (k = 1, ..., p)$$

Simulation study: Settings

- Three noise variance conditions: $\phi_0^2 = 0.5^2, 0.9^2$ and 1.2^2.
- Two sizes of $\mathbf{X}$ conditions: (12×2) and (18×3)
- 200 artificial datasets were created from the normal distribution,

$$\mathbf{X} \sim N(\mathbf{0}, \mathbf{I}).$$

- Distance matrix $\mathbf{D}$ is calculated from $\mathbf{X}$.
- Noise is introduced,

$$\delta_{ij} \sim N(d_{ij}, \phi_0^2),$$

to generate "observed" dataset $\mathbf{\Delta} = \{\delta_{ij}\}$.

Simulation study: Settings

- Hyperpriors and initial values were set following Oh & Raftery (2001).
- 10,000 MCMC samples were used for estimation after 3,000 burn-in.
- For Method #1, approximate mode is used as an point estimate. For other methods, posterior means are used as point estimates.
- As an evaluation measure, MSE was calculated for each point estimation after centering and Procrustes rotation to the true configuration.

Simulation study: Results

- Mean of MSEs ($\mathbf{X}$: 12×2)

	Method #1	Method #2	Method #3	Proposed
$\lambda_0 = 0.5^2$	1.845	1.522	1.503	1.451
$\lambda_0 = 0.9^2$	6.904	5.184	5.200	5.099
$\lambda_0 = 1.2^2$	11.541	8.218	8.265	7.918

- Mean of MSEs ($\mathbf{X}$: 18×3)

	Method #1	Method #2	Method #3	Proposed
$\lambda_0 = 0.5^2$	5.196	3.872	3.806	3.766
$\lambda_0 = 0.9^2$	16.627	11.568	11.538	11.184
$\lambda_0 = 1.2^2$	28.938	18.678	18.896	18.362

- Proposed method performed the best.

Summary & Discussion

- We proposed a new post-processing approach for rotational indeterminacy problem in MCMC estimation.
- Proposed method best recovered the original configuration in simulation study.
- The proposed method should also be applicable to other models with rotational indeterminacy.
- Further studies on the related models are desired.

Thank you very much for your patience.